REMARKS/ARGUMENTS

Favorable reconsideration of this application is respectfully requested.

Minor formal changes are made to the specification and Abstract.

Claims 1-3, 13-15 and 17 are present in this application, claims 4-12 and 16 being canceled by way of the present amendment. Claims 1-11, 16 and 17 are rejected under 35 U.S.C. § 102(e) over U.S. 6,785,649 (Hoory et al.) and claims 12-15 are rejected under 35 U.S.C. § 103(a) over Horry et al. in view of U.S. 6,665,641 (Coorman et al.).

Claims 1-3, 13-15 and 17 are amended by way of the present amendment. The amended claims are supported by the specification and thus no question of introduction of new matter is believed to be raised. For example, amended claim 1 is believed to be supported by page 54, line 25 – page 55, line 14 and page 62, line 23 – page 63, line 18.

The present invention is directed to a variable voice rate apparatus and method. In the variable voice rate apparatus in claim 1, a reproduction information generation unit is included which is configured to generate, as reproduction information concerning reproduction control of the voice for each of linguistic units, information indicating a probability with which preset ones of the linguistic units are combined in a preset order. The linguistic units are produced by a division information generation unit by dividing text data from a text data generation unit. A non-limiting example of such an apparatus is described with relation to Figures 23-25 where the controller classifies the text data into combinations (1-26) and determines the probability of each of the combinations, referring to statistic priority information. The voice reproduction controller selects combinations of the units having a probability lower than a preset value and controls reproduction of the voice data corresponding to the selected combinations. The reproduced voice data can be reproduced at a rate corresponding to the combinations. For example, in one non-limiting manner the apparatus can reproduce common expressions that are more easily understood at a higher rate

than uncommon expressions, allowing the voice data to be reproduced sufficiently understandably while in a shorter time.

Turning to the 35 U.S.C. § 103 rejections, Hoory et al. discloses a text formatting method for converting speech into text and vice versa. Spoken input is received at a microphone 12 which converts the speech into an electrical audio signal. Processor 16 receives the audio signal, performs speech analysis, and generates text corresponding to the speech received from microphone 12. Parameters of the speech are determined, such as word rate, word volume and word pitch. Based on an analysis of the rate, volume and pitch, integral values are assigned and a mapping occurs between the integral values and the format characteristics of the text which is generated, as illustrated in Table 1 in column 7. Figure 3 illustrates examples of the text produced according to the method taught by Hoory et al.

However, there is no disclosure, or any recognition whatsoever, of an apparatus having a reproduction generation unit configured to generate, as reproduction information concerning reproduction control of the voice for each of linguistic units, information indicating a probability with which preset ones of the linguistic units are combined in a preset order, and a voice reproduction controller which selects, from the linguistic units, combinations of the linguistic units each having a probability lower than a preset value and controls reproduction of the voice data corresponding to the selected combinations, as recited in Claims 1 and 2. There is also no disclosure or suggestion of an apparatus having reproduction generation unit configured to generate, as reproduction information concerning reproduction control of the voice for each of linguistic units, information indicating a probability with which preset ones of the linguistic units are combined in a preset order, and a voice signal selection unit configured to select from the linguistic units combinations having a probability lower than a preset value as recited in claim 3. Accordingly, claims 1-3 are patentable over Hoory et al.

Hoory et al. also does not disclose or suggest any method of controlling a reproduction rate of voice including a step of generating, as reproduction information concerning reproduction control of the voice for each of linguistic units, information indicating a probability with which preset ones of the linguistic units are combined in a preset order, and controlling reproduction of voice data corresponding to the selected combinations, as recited in claim 17. As discussed above, Hoory et al. contains no disclosure or any recognition whatsoever of generating information indicating a probability or controlling reproduction of voice data corresponding to selected combinations, as recited in the method of claim 17. Claim 17 is also patentable over Hoory et al.

Coorman et al. describes a speech synthesis system where a text processor 101 converts text into an input phonetic data sequence which may be converted by a target generator 111 into a multi-layer internal data sequence which can include phonetic descriptors, symbolic descriptors and prosodic descriptors. A waveform selector 131 retrieves from the speech unit database 141 descriptors of candidate speech units that can be concatenated into the target utterance specified by the XPT transcription (see column 9, lines 1-29). The waveform selector determines which candidate speech units can be concatenated without causing disturbing quality degradations. Successive candidate speech units are evaluated by waveform selector 131 according to a quality degradation cost function (column 9, lines 38-44). Candidate-to-candidate matching uses frame-based information such as energy, pitch and spectral information to determine how well the candidates can be joined together. Using dynamic programming, the best sequence of candidate speech units is selected for output to the speech waveform concatenator 151 (column 9, lines 45-50).

As described in column 11, beginning at line 19, the waveform selector uses dynamic programming to find the best sequence of diphones such that the database diphones in the best sequence are similar to the target diphones in terms of stress, position, context, etc. and

the database diphones in the best sequence can be joined together with low concatenation artifacts. Cost functions are used to score the suitability of each candidate diphone to be used to synthesize a particular target and to score the joinability of the diphones.

The system of Coorman et al. does not disclose the reproduction information generation unit in claims 1-3 which generates information indicating a probability with which preset ones of the linguistic units are combined in a preset order. Coorman et al. further does not suggest a voice reproduction controller which selects combinations of the units each having a probability lower than a preset value as recited in claims 1 and 2, or the voice signal selector of claim 3. Coorman et al. is directed to determining the ideal fit between two diphones to provide the best match as to certain parameters. Coorman et al. does not have any structure to generate, for each linguistic unit, information indicating a probability with which preset ones of the linguistic units are combined, nor any apparatus which would select units each having a probability lower than a preset value. Coorman et al. simply does not assign probabilities to linguistic units and then select those units having a probability lower than a preset value. Coorman et al. looks at the desirability and suitability of joining two diphones, a different approach from assigning probabilities to each linguistic unit and selecting those having a probability lower than a preset value.

Accordingly, it is clear that Coorman et al. does not disclose any apparatus having a reproduction information generating unit and a voice reproduction controller as recited in each of claims 1 and 2, or a reproduction information generating unit and the voice signal selection unit of claim 3. Thus, even if Coorman et al. could be combined with Hoory et al., the combination would still fail to disclose or suggest the apparatus as recited in each of claims 1-3.

Claim 17 recites a method including a step of generating information including a probability with which preset ones of the linguistic units are combined in a preset order for

each of the linguistic units, and a step of selecting combinations of the linguistic units each having a probability lower than a preset value, based upon the production information and the division information. As is apparent from the discussion of Coorman et al. above, there is no disclosure of any method where reproduction information is generated for each of the linguistic units indicating a probability with which preset ones of the linguistic units are combined in a preset order, and selecting combinations of the linguistic units each having a probability lower than a preset value. Coorman et al. looks at the desirability of joining two diphones and contains no suggestion of assigning reproduction information to each of the linguistic units and then selecting combinations of these linguistic units each having a probability lower than a preset value, based on the reproduction information and the division information.

Accordingly, claim 17 is also patentably distinguishable over a combination of Hoory et al. and Coorman et al. since, even if Coorman et al. could be combined with Hoory et al., the combination would fail to suggest the method of claim 17.

It is respectfully submitted that the present application is in condition for allowance. A favorable decision to that effect is respectfully requested.

Respectfully submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.

Carl E. Schlier

Eckhard H. Kuesters
Registration No. 28,870

Carl E. Schlier
Registration No. 34,426
Attorneys of Record

Customer Number
22850

Tel: (703) 413-3000
Fax: (703) 413 -2220
(OSMMN 08/07)

I:\ATTY\CES\247098US-AM DUE 9-21-07.DOC